Homework 9

Christophe Hunt

April 1, 2017

Contents

This week, we'll empirically verify Central Limit Theorem. We'll write code to run a small simulation on some distributions and verify that the results match what we expect from Central Limit Theorem. Please use R markdown to capture all your experiments and code. Please submit your Rmd file with your name as the filename.

1 (1) Produce a sample of random variable

First write a function that will produce a sample of random variable that is distributed as follows:

$f(x) = x, 0 \leq x \leq 1$
$f(x) = 2 - x, 1 < x \leq 2$

That is, when your function is called, it will return a random variable between 0 and 2 that is distributed according to the above PDF. Please note that this is not the same as writing a function and sampling uniformly from it. In the online session this week, I'll cover Sampling techniques. You will find it useful when you do the assignment for this week. In addition, as usual, there are one-liners in R that will give you samples from a function. We'll cover both of these approaches in the online session.

```
random <- function(){
          num <- runif(1, min = 0, max = 2)
          if(num > 1 && num <= 2){
             return(2-num)
            } else {
              return(num)
            }
          }
random()
```

```
## [1] 0.650578
```

2 (2) Produce a sample of random variable

Now, write a function that will produce a sample of random variable that is distributed as follows:

$f(x) = 1 - x, 0 \leq x \leq 1$
$f(x) = x - 1, 1 < x \leq 2$

```
random2 <- function(){
          num <- runif(1, min = 0, max = 2)
          if(num > 1 && num <= 2){
             return(num-1)
            } else {
              return(1-num)
            }
          }
random2()
```

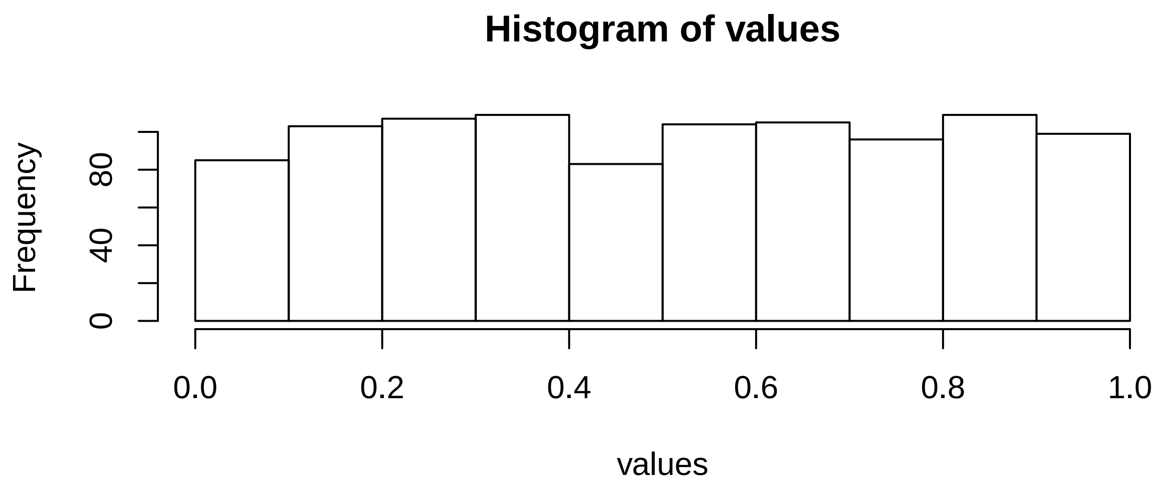
```
## [1] 0.5838077
```

3 (3) Draw 1000 samples

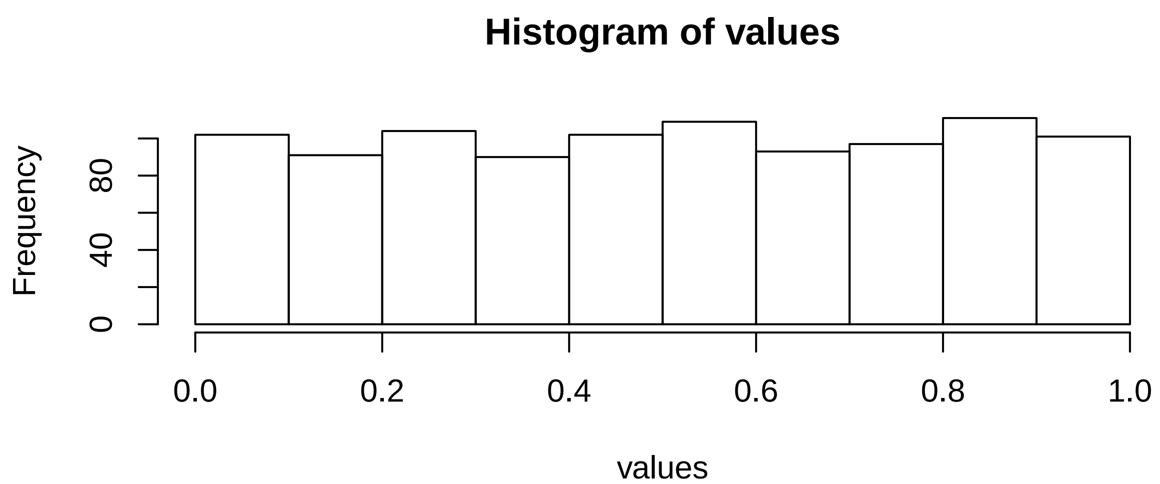
Draw 1000 samples (call your function 1000 times each) from each of the above two distributions and plot the resulting histograms. You should have one histogram for each PDF. See that it matches your understanding of these PDFs.

> The histograms do match my understanding of the PDFs as I expect the values to be between 0 and 1 for both PDFs.

```
values <- replicate(1000, random())
hist(values)
```




```
values <- replicate(1000, random2())
hist(values)
```



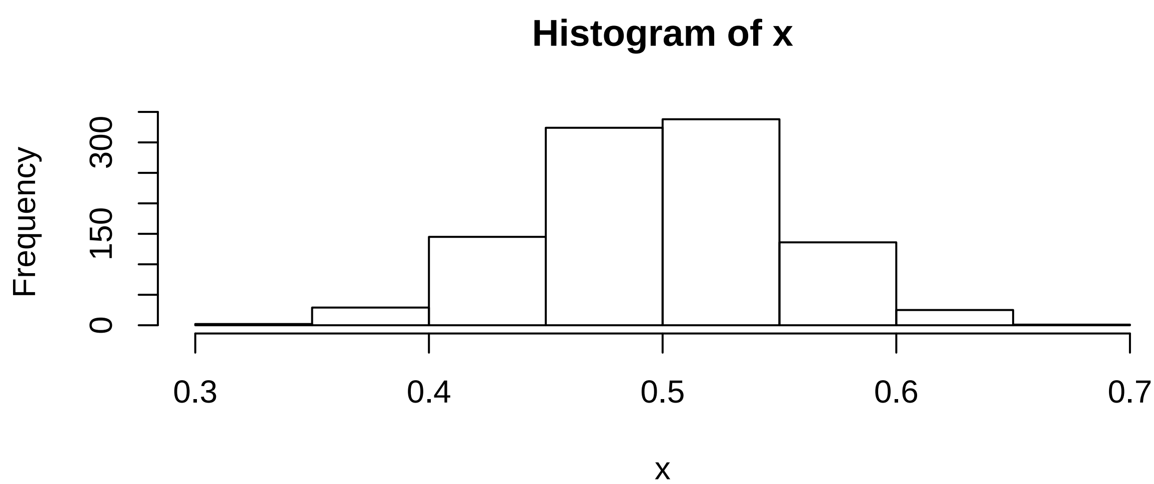
4 (4) Function to take means and generate histogram

Now, write a program that will take a sample set size n as a parameter and the PDF as the second parameter, and perform 1000 iterations where it samples from the PDF, each time taking n samples and computes the mean of these n samples. It then plots a histogram of these 1000 means that it computes.

```
means.hist <- function(n, PDF){
          x <- NULL
          for (i in 1:1000){
           z <- mean(replicate(n, PDF()))
           x <- rbind(x, z)
          }
          hist(x)
}
```

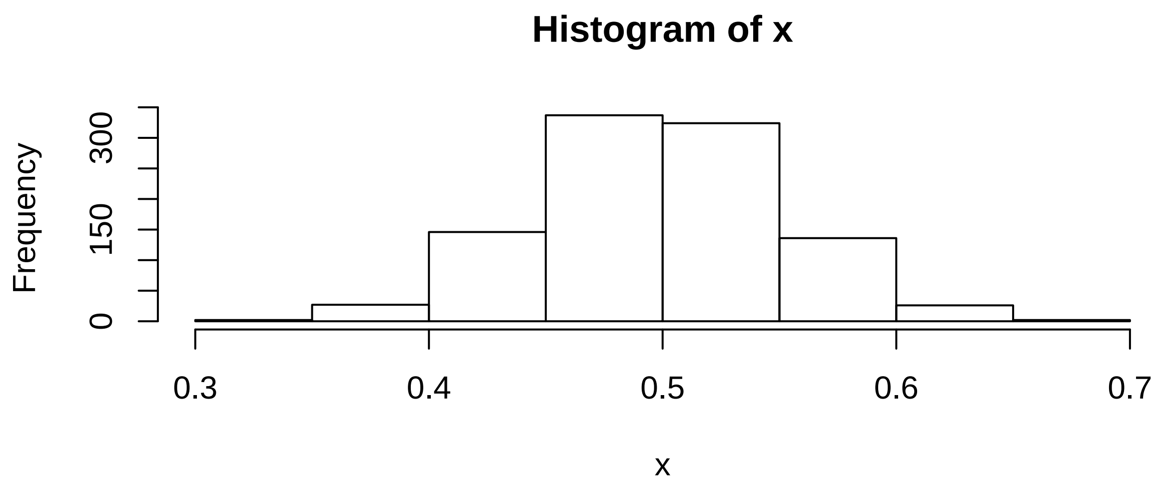
4.1 Histogram of first random function

```
means.hist(30, PDF = random)
```



4.2 Histogram of second random function

```
means.hist(30, PDF = random2)
```

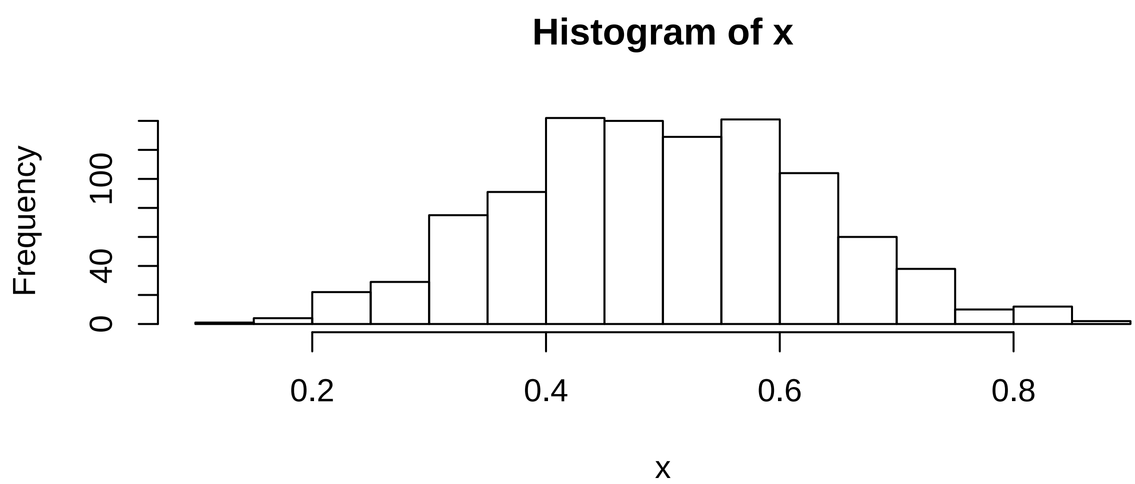


5 (5) Empirically verifying the Central Limit Theorem

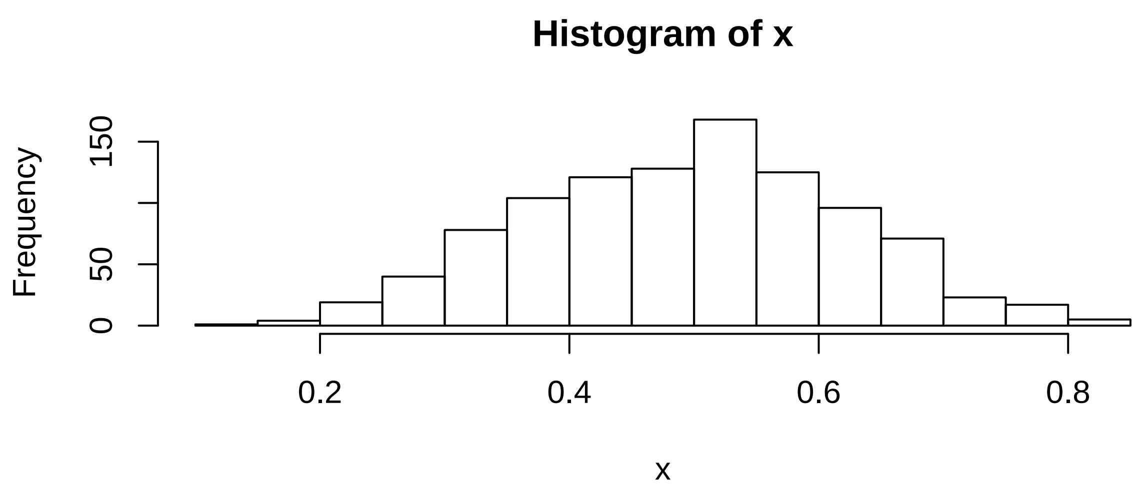
Verify that as you set n to something like 10 or 20, each of the two PDFs produce normally distributed mean of samples, empirically verifying the Central Limit Theorem. Please play around with various values of n and you'll see that even for reasonably small sample sizes such as 10, Central Limit Theorem holds.

> It is true that the central limit theorem holds at very low values. Even at only 5 iterations we can see the normal distribution appearing.

```
means.hist(5, PDF = random)
```

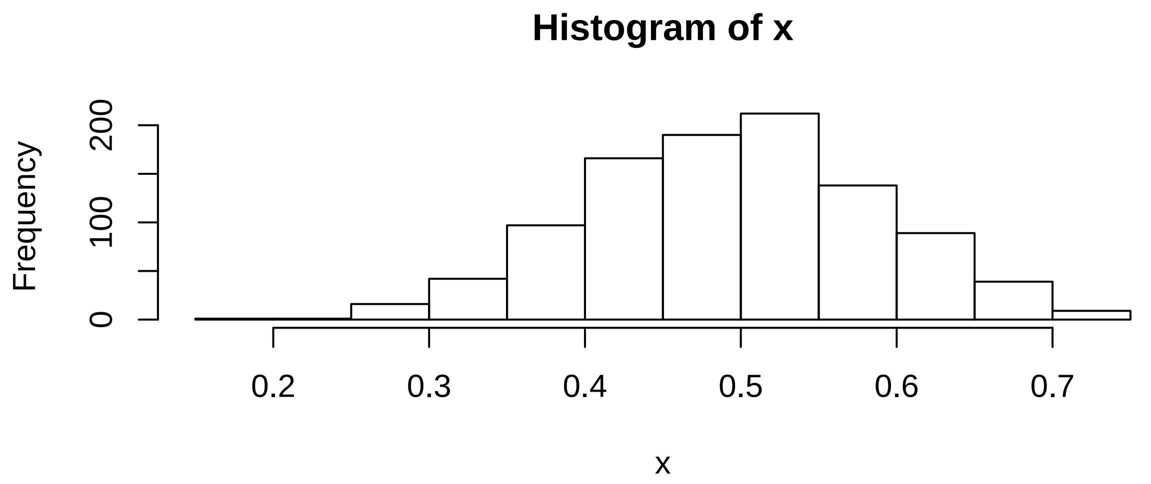



```
means.hist(5, PDF = random2)
```



> At the n = 10 suggestion the central limit theorem becomes very clear.

```
means.hist(10, PDF = random)
```




```
means.hist(10, PDF = random2)
```